

# REVUE DE STATISTIQUE APPLIQUÉE

J. LELLOUCH

## **L'application de l'analyse séquentielle aux essais thérapeutiques**

*Revue de statistique appliquée*, tome 17, n° 1 (1969), p. 41-68

[http://www.numdam.org/item?id=RSA\\_1969\\_\\_17\\_1\\_41\\_0](http://www.numdam.org/item?id=RSA_1969__17_1_41_0)

© Société française de statistique, 1969, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# L'APPLICATION DE L'ANALYSE SÉQUENTIELLE AUX ESSAIS THÉRAPEUTIQUES

J. LELLOUCH

Unité de Recherches Statistiques de l'INSERM

## INTRODUCTION

Dès la parution des travaux de WALD, on a pensé à adapter les méthodes séquentielles au problème de la comparaison de deux traitements A et B dans un essai thérapeutique (E.T.). En effet, quand on décide de réaliser un E.T. sur un nombre N de malades, il est exceptionnel de disposer d'emblée de la totalité de ces sujets et de les traiter simultanément. Les malades entrent dans l'essai l'un après l'autre, pendant un temps T qui peut durer plusieurs mois. Si le critère utilisé pour comparer les deux thérapeutiques est disponible après un temps depuis le début du traitement qui est petit par rapport à T, on commence à avoir des réponses à un moment où peu de malades sont encore entrés dans l'essai. Si les différences entre les traitements à comparer sont importantes, ou au contraire tout à fait négligeable il paraît souhaitable de s'en apercevoir et d'arrêter l'essai le plus tôt possible. On réalisera ainsi un gain sur les deux paramètres limitant les E.T. : le nombre de sujets et son corollaire, la durée de l'essai.

Si quelques E.T. ont effectivement été planifiés selon le schéma primitif de WALD, celui-ci a été assez rapidement abandonné pour deux raisons essentielles ; la première est qu'il est "ouvert" : si le nombre moyen de sujets nécessaire est inférieur à celui d'une analyse à effectif fixé, il se peut que dans un essai déterminé, il soit très nettement supérieur ; la deuxième est qu'il correspond à des problèmes à deux réponses. Aussi ARMITAGE [5] et quelques autres ont-ils proposé des schémas fermés et à trois conclusions ( $A < B$  ;  $A = B$  ;  $A > B$ ). La démarche adoptée dans tous ces tests est la démarche classique de NEYMAN et PEARSON ; elle est depuis peu sévèrement critiquée par quelques théoriciens, dont COLTON [7], ANSCOMBE [3], CORNFIELD [8], qui préfèrent un point de vue Bayésien.

On s'est d'autre part rendu compte que la formulation "Comparaison de deux traitements A et B", est absolument insuffisante, en ce sens qu'elle recouvre divers problèmes, qu'on peut en général grouper autour de deux attitudes extrêmes qualifiées d'"explicative" et de "pragmatique" (19,20) .

## L'ATTITUDE EXPLICATIVE ET L'ATTITUDE PRAGMATIQUE

En attitude explicative, il est intéressant du point de vue biologique, par exemple pour vérifier une hypothèse, de savoir si une différence existe ou non entre les deux traitements ; il est grave de la proclamer alors qu'en réalité elle est nulle (risque  $\alpha$  de première espèce); il est, de même ennuyeux

de ne pas déceler une différence qui existe réellement (risque  $\beta$  de deuxième espèce). C'est l'attitude du chercheur de laboratoire, pour qui le risque de première espèce est l'élément prépondérant ; c'est celle exposée dans les cours élémentaires de méthodologie statistique.

Deux éventualités sont cependant à considérer en attitude explicative : dans la situation bilatérale, on teste que la différence  $B-A$  est nulle, contre l'hypothèse alternative qu'elle est soit positive, soit négative ; cette situation se rencontre quand on ignore de quel signe est la différence, si elle existe ; trois conclusions sont possibles ( $A < B$  ;  $A$  non différent significativement de  $B$  ;  $A > B$ ). Dans la situation unilatérale, l'hypothèse alternative est que la différence est d'un signe déterminé, par exemple positive, la réponse  $B-A < 0$  étant soit inintéressante, soit exclue ; deux conclusions sont alors possibles ( $B > A$ ,  $B$  non supérieur significativement à  $A$ ). Nous pensons, toutefois que cette situation unilatérale est rare en attitude explicative ; nous reviendrons plus longuement sur ce point au paragraphe III.

En attitude pragmatique, la situation est toute différente ; il s'agit de choisir le meilleur des deux traitements ; il n'est aucunement gênant d'adopter  $A$  ou  $B$  si  $A = B$  (étant entendu que l'appréciation des résultats repose sur un bilan global). Par contre, il est particulièrement ennuyeux d'adopter un traitement, si c'est le moins bon (risque  $\gamma$  de troisième espèce). C'est l'attitude de celui qui doit prendre une décision. Deux conclusions sont possibles (adopter  $A$  ou adopter  $B$ ).

La distinction entre les deux attitudes est en fait bien plus fondamentale qu'elle ne paraît au premier abord, car elle intervient de façon essentielle dans la définition précise des traitements à comparer, des malades à faire entrer dans l'essai, des critères de jugement permettant la comparaison. Ces points sont complètement étudiés par SCHWARTZ et FLAMANT (18).

Disons seulement ici, en ce qui concerne les critères de jugement, qu'en attitude explicative, on considère habituellement un petit nombre de critères (voire un seul) ayant une signification biologique, qu'on étudie séparément. Il se peut donc que des différences existent pour certains critères et pas pour d'autres.

En attitude pragmatique les critères sont généralement nombreux, mais ils ne peuvent être considérés isolément, puisqu'à la fin de l'essai, une seule décision doit être prise (ou  $A$  ou  $B$ ). On peut les combiner en un critère unique par une comparaison pondérée, les poids étant fonction de l'importance pratique attribuée à chacun d'eux, ce critère final étant soit quantitatif, soit transformé en critère qualitatif à deux classes que nous appellerons  $+$  et  $-$ .

Les deux traitements sont alors "équivalents" ou "symétriques" (on n'a pas de raison de choisir l'un plutôt que l'autre) si les moyennes ou les pourcentages de  $+$  de ce critère unique sont égaux.

Il se peut cependant qu'un des critères joue un rôle tellement important qu'il est difficile de le combiner aux autres (c'est, par exemple le cas du critère survie). On peut alors axer la comparaison sur ce seul critère privilégié, mais décider que  $B$  est équivalent à  $A$  s'il le surpasse d'une certaine quantité  $h$ , résumant ses avantages et inconvénients relatifs pour la totalité des critères accessoires. On peut aussi, ne pas chercher à définir exactement le "point d'équivalence", mais exprimer la dissymétrie des

deux traitements, par un choix convenable des risques d'erreur consentis par le médecin pour différentes valeurs de la différence B-A.

#### ORGANISATION DE L'ESSAI

Que le critère soit quantitatif ou qualitatif, les sujets sont associés par paires, l'un recevant A et l'autre B, de telle manière qu'on puisse faire le point après chaque paire (voir plus loin, la constitution de ces paires).

Si le critère de jugement est quantitatif, soit  $x$ , on considère pour chaque paire la différence  $d = x_B - x_A$ .

La statistique intervenant dans les tests est  $\Sigma d$  (ou une quantité voisine) ; on peut donc représenter graphiquement les résultats observés en dessinant le "chemin" donnant la variation de  $\Sigma d$  en fonction de  $n$ , nombre de paires entrées dans l'essai. On arrête celui-ci dès que le chemin rencontre certaines frontières appropriées.

Si le critère est qualitatif à deux classes, on élimine, comme dans le test de comparaison de pourcentages de deux séries appariées, les paires ++ et -- ; pour les paires restantes, dites "utiles", la comparaison est exprimée par la différence  $d$ , cotée + 1 si la réponse positive a été obtenue avec B, - 1 si elle a été obtenue avec A. La statistique intervenant dans les tests est encore  $\Sigma d$ , et le "chemin" est constitué de segments orientés à 45° vers le haut ou vers le bas.

Si  $P_A$  et  $P_B$  désignent les pourcentages de + avec A et B,  $d$  est égale à + 1 avec la probabilité

$$\theta_B = \frac{P_B(1 - P_A)}{P_B(1 - P_A) + P_A(1 - P_B)}$$

et -1 avec la probabilité  $\theta_A = 1 - \theta_B$ . Sa moyenne est  $2\theta_B - 1$ , et sa variance  $4\theta_B\theta_A$ .

On voit que  $\Sigma d$  est le nombre de paires où le résultat avec B s'est révélé meilleur que celui avec A diminué du nombre de paires ayant la propriété inverse ; ce point de vue est très utile dans les cas où il est difficile de classer objectivement les résultats en + ou -, mais où il est plus facile de les comparer qualitativement : par exemple les deux traitements étant administrés au même sujet (méthode du sujet pris comme son propre témoin), celui-ci pourra dire qu'il a préféré A à B. Bien entendu avec cette notion de préférence, on ne peut plus parler de  $P_A$  et  $P_B$ .

Reste à dire un mot sur la constitution des paires : dans l'analyse à effectif fixé, l'appariement a pour but l'augmentation de la puissance des tests ; en analyse séquentielle, il est nécessaire pour que la théorie reste aisée. Il est bien entendu avantageux de prendre comme éléments de la paire deux sujets aussi semblables que possible et en particulier de choisir le sujet comme son propre témoin, si c'est possible. Sinon, on peut se baser sur des critères jugés importants, tels que le sexe, l'âge, la gravité de la maladie. Cependant on risque alors de ne pas disposer du deuxième élément de la paire avant un long temps. Aussi la plupart du temps, construit-on les paires uniquement selon l'ordre d'entrée dans l'essai.

Dans les pages qui suivent on décrira une gamme assez large de tests séquentiels qui ont été proposés pour les E.T.. Certains ont essentiellement un intérêt théorique, d'autres ont été effectivement utilisés lors de la conduite de tels essais, ou au moins dans le domaine industriel. Ils sont tous basés sur la théorie des tests de NEYMAN et PEARSON. Les procédés (Bayesiens ou autres) qui dérivent de la théorie de la décision sont actuellement en plein développement. Ils seront exposés ultérieurement.

## I - PROBLEMES A DEUX CONCLUSIONS

(Attitude Pragmatique. Attitude Explicative, situation unilatérale).

A. Critère quantitatif (1), la variance  $\sigma^2$  de la différence  $d = x_B - x_A$  est connue

Les exigences de l'expérimentateur peuvent être consignées dans le tableau général ci-dessous :

$\mu_B - \mu_A$	$\Delta_0(H_0)$	$\Delta_1(H_1)$
Conclusion souhaitée	Adopter A	Adopter B
Risque consenti	$\gamma_0$	$\gamma_1$

En attitude pragmatique, cas symétrique, on choisira  $\Delta_1$  et  $\Delta_0$  symétrique par rapport à 0, et des risques  $\gamma_0$  et  $\gamma_1$  égaux. Si le point d'équivalence est h, on choisira  $\Delta_0 = h + \Delta$ ,  $\Delta_1 = h - \Delta$ ;  $\gamma_1$  et  $\gamma_0$  égaux. En attitude explicative, on prend  $\Delta_0 = 0$ ;  $\Delta_1 = \Delta$ ,  $\gamma_0 = \alpha$ , risque de première espèce,  $\gamma_1 = \beta$ , risque de deuxième espèce, les conclusions souhaitées étant dans ce cas, B non significativement supérieur à A et B supérieur à A.

### 1. Test ouvert de WALD [23]

On sait que si  $L_1$  et  $L_0$  sont les vraisemblances de l'échantillon sous  $\Delta_1$  et  $\Delta_0$ , on continue l'échantillonnage tant que

$$l_0 < \frac{L_1}{L_0} < l_1,$$

et on termine par la conclusion appropriée dès qu'il n'en est plus ainsi,  $l_0$  et  $l_1$  étant des constantes qui donnent aux risques les valeurs souhaitées. Des valeurs approchées pour  $l_0$  et  $l_1$  sont

$$\frac{\gamma_1}{1 - \gamma_0} \quad \text{et} \quad \frac{1 - \gamma_1}{\gamma_0}$$

Ces valeurs peuvent s'obtenir de plusieurs manières. La façon la plus intéressante est d'approcher le chemin discret où les accroissements de  $\Sigma d$  ne se produisent que pour des valeurs entières des abscisses par un chemin continu, représentant un processus de WIENER.

-----  
(1) dans la suite d est supposée distribuée normalement.

(On rappelle que si la variable X dépendant du temps t, suit un processus de WIENER de moyenne  $\mu$  et de variance  $\sigma^2$ , l'accroissement dX pendant le temps dt est une variable normale de moyenne  $\mu dt$  et de variance  $\sigma^2 dt$ , les accroissements pendant des intervalles successifs étant indépendants en probabilité).

Ceci conduit à deux frontières qui sont les droites parallèles (le schéma est donc "ouvert") de pente

$$b = \frac{\Delta_0 + \Delta_1}{2}$$

et d'ordonnées à l'origine

$$c_+ = \frac{\sigma^2}{\Delta_1 - \Delta_0} \log \frac{1 - \gamma_1}{\gamma_0}$$

$$c_- = \frac{\sigma^2}{\Delta_1 - \Delta_0} \log \frac{\gamma_1}{1 - \gamma_0}$$

Un cas particulier intéressant est celui où les 2 traitements jouent un rôle symétrique : les droites sont alors horizontales et

$$c_+ = c_- = \frac{\sigma^2}{2\Delta} \log \frac{1 - \gamma}{\gamma}$$

Les propriétés du test de WALD sont classiques :

- la probabilité est un que l'échantillonnage se termine.

- le nombre de paires nécessaire, n, pour aboutir à une conclusion est une variable aléatoire, dont il est facile de calculer la moyenne E(n) en fonction de la différence  $\mu_B - \mu_A = \mu$ . E(n) passe par un maximum pour une valeur située entre  $\Delta_0$  et  $\Delta_1$ . Dans le cas pragmatique symétrique, le maximum est atteint pour une différence nulle ; s'il y a un handicap h, le maximum à lieu pour h. Ceci est tout à fait satisfaisant du point de vue éthique (il faut d'autant moins de sujets que les traitements diffèrent plus). E(n) est très généralement inférieur au nombre nécessaire dans une analyse à effectif fixé ;

- le test de WALD est celui qui minimise E(n) pour des valeurs des différences égales à  $\Delta_0$  et  $\Delta_1$ .

- on peut également déterminer la courbe d'efficacité du test ; on vérifie immédiatement dans le cas pragmatique, que la probabilité d'accepter B ou A est bien 1/2 quand les deux traitements sont "équivalents".

Du point de vue de la moyenne E(n), le test de WALD paraît satisfaisant : mais la variance de n étant relativement importante, il se peut que dans un essai particulier, on ait besoin d'un nombre très élevé de paires. Il y a donc une très grande difficulté à prévoir, même approximativement, combien de temps durera un essai. Cet inconvénient qui est dû au caractère ouvert du test, est évidemment une gêne pratique considérable, qui a fait abandonner presque complètement en médecine ce type de schéma.

## 2. Tests fermés

Une façon de remédier à l'inconvénient des tests ouverts (possibilité d'essais très longs) est de décider que de toutes façons, on ne prendra pas plus de  $N$  paires,  $N$  étant un nombre fixé à l'avance.

(a) avec un système de droites parallèles on obtiendrait un schéma tel que celui représenté sur la figure 1. La règle de décision étant la suivante : accepter B dès que le chemin rencontre la frontière  $D_1 M_1 M$  ; accepter A dès que le chemin rencontre la frontière  $D_0 M_0 M$ . Bien entendu les droites  $D_0$  et  $D_0$  ne sont plus celles de WALD, et il y a une infinité de systèmes  $(D_1, D_0, M)$  satisfaisant aux exigences de l'expérimentateur, parmi lesquels il est bien entendu extrêmement difficile de choisir.

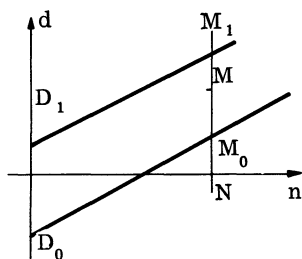


Fig. 1

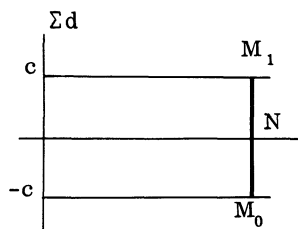


Fig. 2

ANDERSON [1] en assimilant le processus discret à un processus continu, calcule pour un système  $(D_1, D_0, M, N)$ , la probabilité d'accepter le traitement B (ou A) et le nombre moyen de paires nécessaire à une conclusion, en fonction de la différence réelle  $\mu_B - \mu_A$ . Son travail a un intérêt théorique considérable, mais les formules qu'il donne sont extraordinairement compliquées, et pour l'instant de peu d'intérêt pratique.

Un seul cas a été étudié du point de vue numérique par GENIZI [12] : c'est le cas symétrique représenté sur la figure 2.

On voit que le test de WALD et la méthode à effectif fixé sont des cas particuliers de ce modèle tronqué : la méthode séquentielle correspond à  $N$  infini et

$$c = \frac{\sigma^2}{2\Delta} \log \frac{1-\gamma}{\gamma}$$

tandis que la méthode classique correspond à

$$N = (\varepsilon_{2\gamma})^2 \frac{\sigma^2}{\Delta^2}$$

où  $\varepsilon_{2\gamma}$  est la valeur de l'écart réduit relatif à un risque  $\gamma$ , et  $c$  infini.

Pour un  $N$  donné, il existe un  $c$  et un seul, qui donne aux risques d'erreur les valeurs fixées ; mais on peut prendre pour  $N$  toute une série de valeurs ; autrement dit il existe toute une famille de modèles possibles, compris entre les 2 extrêmes : méthode séquentielle, méthode classique.

La table donne pour différentes valeurs de  $N$ , les  $c$  correspondants. Elle a été construite pour  $\sigma^2=1$ , et la "limite de sécurité" ( $\Delta = 0,1$  ;  $\gamma = 0,05$ ).

Les valeurs sont approximatives.

N	271	280	300	350	400	$\infty$
c	$\infty$	20,0	17,8	15,9	15,3	14,7

Nota : Si  $\sigma^2 \neq 1$  et  $\Delta \neq 0,1$  on doit multiplier les valeurs lues dans la table par  $\left(\frac{\sigma}{10\Delta}\right)^2$  pour obtenir N, et par  $\frac{\sigma^2}{10\Delta}$  pour obtenir c.

La plus petite valeur de N est bien celle donnée par  $(\varepsilon_{2,\gamma})^2 \frac{\sigma^2}{\Delta^2}$ ; N est donc supérieur au nombre nécessaire pour une analyse classique. Dans un essai particulier on peut perdre en nombre de sujets par rapport à la méthode classique, mais cette perte est nécessairement limitée, contrairement à ce qui peut se passer avec les modèles "ouverts"; d'autre part, on a toutes chances d'y gagner.

Il reste qu'en face d'un problème déterminé, il faut choisir la combinaison (c, N) à utiliser; on peut prendre pour N le nombre maximum de paires qu'il est possible de faire entrer dans l'essai et en déduire le c correspondant.

Il convient de remarquer qu'ANDERSON considère également des systèmes de droites non parallèles, en particulier des systèmes de droites concourantes, conduisant donc à un test nécessairement fermé. Ces tests paraissent très intéressants, en ce qui concerne les valeurs de  $E(n)$ , pour des différences comprises entre  $\Delta_0$  et  $\Delta_1$ , mais beaucoup reste à faire du point de vue numérique.

b) Une autre façon de procéder est due à ARMITAGE [4].

On conserve la frontière supérieure du test de WALD ( $\Sigma d = bn + c$ ) et on remplace la frontière inférieure par une verticale d'abscisse N, (d'où le nom de tests "limités" qu'on donne à ces procédés), N étant déterminé de telle façon que les risques fixés soient préservés (Fig. 3).

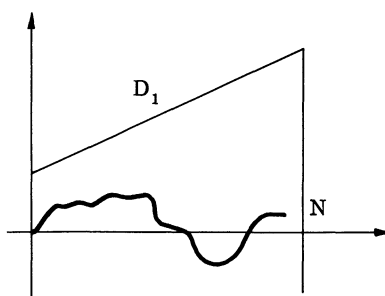


Fig. 3

La probabilité  $P(\mu, N)$  que le chemin coupe  $D_1$  avant la N paire, lorsque la différence est  $\mu$ , est donnée par la formule (toujours par assimilation à un processus continu):



$$P(\mu, N) = 1 - \Phi \left\{ \frac{-(\mu - b)N + c_+}{\sigma\sqrt{N}} \right\} + \exp \left\{ \frac{2c_+(\mu - b)}{\sigma^2} \right\} \Phi \left\{ \frac{-(\mu - b)N - c_+}{\sigma\sqrt{N}} \right\}$$

où  $\Phi$  est l'intégrale de la loi normale réduite.

On peut obtenir cette formule, comme cas particulier d'une formule générale d'ANDERSON ( $D_0$  repoussée à l'infini,  $M_1$  et  $M$  confondus) ; on peut également la démontrer directement de plusieurs façons ; la plus élémentaire est sans doute celle indiquée par FELLER (11, chap. XIV).

On exige

$$P(\Delta_0, N) = \gamma_0$$

$$P(\Delta_1, N) = 1 - \gamma_1$$

d'où deux équations pour calculer  $N$ . Il est cependant facile de voir que ces 2 équations sont compatibles (ceci résultant du fait que le rapport des vraisemblances  $L_1/L_0$  vaut justement  $\frac{1 - \gamma_1}{\gamma_0}$  sur  $D_1$ ).

On peut donc tirer  $N$  de l'une quelconque des deux équations.

ARMITAGE [4] donne une table permettant de calculer  $N$  pour différentes combinaisons des risques  $\gamma_0$  et  $\gamma_1$ .

Du fait que ce test fait jouer un rôle tout à fait dissymétrique aux deux conclusions il ne paraît intéressant qu'en attitude explicative où on conclut à la différence si  $D_1$  est traversée avant  $N$ , à la non-signification dans le cas contraire.

c) Si dans le test précédent, un chemin est encore loin de la frontière supérieure pour un  $n$  assez proche de  $N$ , il n'a que très peu de chance de la rencontrer, d'où l'idée de SCHNEIDERMAN et ARMITAGE [16] de le modifier pour obtenir des frontières telles que celles représentées sur la figure 4, conduisant à des tests dits "en coin" à cause de la forme des frontières.

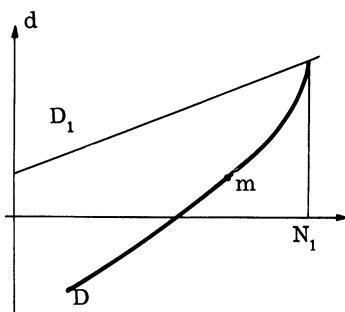


Fig. 4

La droite  $D_1$  est encore la droite supérieure du test de WALD, ou celle du test précédent, la courbe  $D$  étant déterminée pour que les risques fixés soient (au moins approximativement) préservés. Le procédé décrit dans (16) est le suivant :

Supposons que dans le test limité précédent la troncature ait lieu à  $N_1 > N$ . Alors les risques d'erreur ne sont plus évidemment  $\gamma_0$  et  $\gamma_1$ , mais

deviennent  $\gamma'_0$  et  $\gamma'_1$ .  $\gamma'_0$  est bien entendu supérieur à  $\gamma_0$ , disons  $\gamma'_0 = \gamma_0 + \varepsilon$  ( $\varepsilon > 0$ ), et on a, d'après une remarque du paragraphe précédent :

$$\frac{1 - \gamma'_1}{\gamma'_0} = \frac{1 - \gamma_1}{\gamma_0}$$

Définissons D comme le lieu des points m tels que sous  $\Delta_0$  un chemin partant de m ait une probabilité  $\varepsilon'$  de traverser  $D_1$  avant  $N_1$ . Alors la probabilité qu'un chemin partant de l'origine traverse  $D_1$  avant D est sous  $H_0$  :

$$p = \gamma'_0 - \varepsilon'(1 - p)$$

Comme on désire que p soit égal à  $\gamma_0$ , de  $\gamma_0 = \gamma'_0 + \varepsilon - \varepsilon'(1 - \gamma_0)$ , on tire :

$$\varepsilon' = \frac{\varepsilon}{1 - \gamma_0}$$

On peut donc choisir une valeur quelconque de  $\varepsilon'$  : on en déduit  $\varepsilon$ ,  $\gamma'_0 = \gamma_0 + \varepsilon$ , et  $N_1$  par  $P(\Delta_0, N_1) = \gamma'_0$ . Quand à D, on la détermine également à partir de la formule donnant  $P(\mu, N)$ , en ramenant l'origine des coordonnées à m.

Il est difficile de calculer exactement le deuxième risque, mais on peut espérer qu'il est très voisin de la valeur  $\gamma_1$  désirée.

On remarque qu'on a toute une famille de plans possibles correspondant à des valeurs différentes de  $\varepsilon'$ . On peut facilement voir que la valeur minimum de  $\varepsilon'$  est 0, donnant le schéma du paragraphe b, et que sa valeur minimum est

$$\frac{\gamma_0 \gamma_1}{(1 - \gamma_0)(1 - \gamma_1)}$$

donnant le schéma de WALD à 2 parallèles. (Les plans détaillés numériquement en (16) correspondent à  $\varepsilon' = \alpha/100$ ).

On remarque également que ce plan ressemble beaucoup à celui d'ANDERSON [1] constitué de deux droites concourantes. Son intérêt, par rapport au test limité est que  $E(n)$  est bien entendu, nettement plus faible pour B-A voisin de 0. Il ne semble cependant pas avoir été jamais utilisé.

#### B. Le critère est quantitatif, la variance de la différence étant inconnue.

Ce cas, le plus fréquent en pratique, pose un problème nouveau, les hypothèses testées étant en fait composites. Il existe deux méthodes générales de construction de tests séquentiels pour de telles situations : la première due à WALD [23] est connue sous le nom de "théorie des Weight Functions" ; la seconde consiste à travailler sur des fonctions convenables des observations.

Tous les tests "t-séquentiels" concernent des hypothèses sur les valeurs, non des différences absolues, mais des différences rapportées à leur écart-type  $\sigma$ .

$\mu = \mu_B - \mu_A$	$\delta_0 \sigma (H_0)$	$\delta_1 \sigma (H_0)$
Conclusion souhaitée	Adopter A	Adopter B
Risques consentis	$\gamma_0$	$\gamma_1$

Exprimer les différences en termes de  $\sigma$ , peut ne pas paraître toujours satisfaisant dans les problèmes pragmatiques, où ce sont leurs valeurs réelles qui ont un sens concret. Cependant, comme on a généralement avant le début de l'essai une idée de  $\sigma$ , on peut à partir de  $\Delta$  intéressants, obtenir les  $\delta$  à faire intervenir dans la définition du schéma.

### 1. Tests ouverts

#### a. Méthode de WALD [23]

Soit le plan cartésien  $(\sigma, \mu)$ ,  $\theta$  le point courant de ce plan (fig. 5). Si  $\theta \in \omega_0$ , on veut accepter A, si  $\theta \in \omega_1$  on veut accepter B avec les conditions :

$$\gamma_0(\theta) = \Pr(\text{adopter B} / \theta \in \omega_0) \leq \gamma_0$$

$$\gamma_1(\theta) = \Pr(\text{adopter A} / \theta \in \omega_1) \leq \gamma_1$$

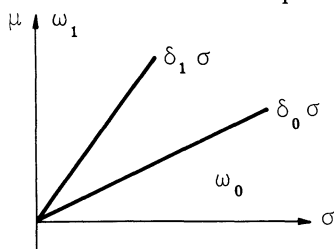


Fig. 5

Soient maintenant  $W_0(\theta)$  et  $W_1(\theta)$  deux distributions de  $\theta$ , telles que

$$\int_{\omega_0} w_0(\theta) d\theta = \int_{\omega_1} w_1(\theta) d\theta = 1.$$

Les  $w_i(\theta)$  sont donc nuls en dehors de  $\omega_i$  ( $i = 0, 1$ ). On peut les considérer comme des distributions a priori de  $\theta$ .

Supposons que nous remplacions le problème du test de l'hypothèse composite  $H_0$  contre l'hypothèse composite  $H_1$ , par le problème de tester que la loi a priori de  $\theta$  est  $W_0$  contre l'alternative qu'elle est  $W_1$ , qui sont 2 hypothèses simples ( $H'_0$  et  $H'_1$ ).

Après la  $n$  observation les distributions de l'échantillon sont :

$$L'_1 = \int_{\omega_1} f(d_1, \theta) \dots f(d_n, \theta) W_1(\theta) d\theta$$

ou  $f(d, \theta)$  est la loi de la variable  $d$ .

Le test de  $W_1$  contre  $W_0$  peut donc s'obtenir à partir d'un test de WALD avec des risques  $\gamma'_0$  et  $\gamma'_1$ , c'est-à-dire en comparant le rapport  $\frac{L'_1}{L'_0}$  aux deux limites  $\frac{\gamma'_1}{1 - \gamma'_0}$  et  $\frac{1 - \gamma'_1}{\gamma'_0}$

On a

$$\int_{\omega_0} W_0(\theta) \gamma_0(\theta) = \gamma'_0$$

$$\int_{\omega_1} W_1(\theta) \gamma_1(\theta) = \gamma'_1$$

Pour n'importe quelles fonctions  $W_0(\theta)$  et  $W_1(\theta)$ ,  $\max. \gamma_0(\theta)$  et  $\max. \gamma_1(\theta)$  sont fonction de  $\gamma'_0$  et  $\gamma'_1$ . On peut donc, théoriquement, choisir  $\gamma'_0$  et  $\gamma'_1$  de telle sorte que ces deux maxima ne dépassent pas les valeurs fixées.

Dans le cas qui nous occupe ici (test t-séquentiel) WALD choisit les fonctions  $W_i$  de la façon suivante :

$W_0$  est concentrée uniformément sur la droite  $\mu = \delta_0\sigma$ , de même  $W_1$  sur la droite  $\mu = \delta_1\sigma$ .

De façon plus précise, en termes de  $\sigma$

$$W_i = \frac{1}{C} \quad 0 < \sigma < C$$

(C est un nombre qu'on fera tendre vers l'infini)

$$W_i = 0 \quad C < \sigma.$$

La distribution de l'échantillon sous  $H'_i$  est alors :

$$\int_0^C \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\sigma^n} \exp. \left\{ -\frac{1}{2\sigma^2} \Sigma(d - \delta_i\sigma)^2 \right\} \frac{d\sigma}{\sigma},$$

de sorte que le test cherché s'obtient en considérant le rapport  $L'_1/L'_0$  où  $L'_i$  vaut

$$\int_0^{+\infty} \frac{1}{\sigma^n} \exp. \left\{ -\frac{1}{2\sigma^2} \Sigma(d - \delta_i\sigma)^2 \right\} d\sigma$$

WALD montre que  $\gamma_0(\theta)$  et  $\gamma_1(\theta)$  sont constants sur les droites  $\mu = d_i\sigma$  de sorte que  $\gamma'_i = \gamma_i$ .

Finalement le test revient à comparer  $L'_1/L'_0$  aux 2 limites habituelles

$$\frac{\gamma_0}{1 - \gamma_0} \quad \text{et} \quad \frac{\gamma_1}{1 - \gamma_1}$$

Une transformation simple de l'intégrale conduit à l'expression équivalente

$$\frac{(n.2)! \exp. \left\{ -\frac{\delta_i^2}{2} (n - u)^2 \right\}}{(\Sigma d^2)^{\frac{n-1}{2}}} \text{Hh}_{n-2}(-\delta u)$$

où u désigne la quantité  $\frac{\Sigma d}{\sqrt{\Sigma d^2}}$ , et  $\text{Hh}_n(x)$  représente l'intégrale

$$\int_0^{\infty} \frac{y^n}{n!} e^{-\frac{1}{2}(y+x)^2} dy.$$

On voit donc que

$$\frac{L'_1}{L'_0} \text{ vaut } \exp \left\{ -\frac{(n-u)^2}{2} (\delta_1^2 - \delta_0^2) \right\} \frac{Hh_{n-2}(-\delta_1 u)}{Hh_{n-2}(-\delta_0 u)}.$$

RUSHTON [14] donne un procédé permettant de calculer le logarithme de ce rapport.

On peut soit le calculer pour chaque n et voir s'il sort des limites,

$$\log \frac{\gamma_1}{1-\gamma_0} \quad \text{et} \quad \log \frac{1-\gamma_1}{\gamma_0}$$

soit tracer les frontières, dans un plan où le chemin représente la variation de  $u = \Sigma d / \sqrt{\Sigma d^2}$  en fonction de n.

On obtient la figure 6.

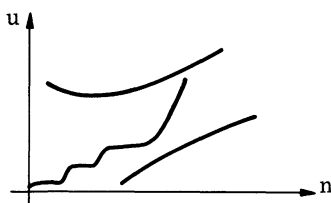


Fig. 6

#### b) Transformation des observations. (RUSHTON [14], BARNARD)

Après chaque nouvelle observation  $d_k$ , on calcule la quantité  $t = \frac{\bar{d}}{s_d^2}$ , où  $s_d^2$  est l'estimation de la variance à  $n-1$  degrés de liberté; et on applique le test de WALD aux observations transformées  $t$ , c'est-à-dire qu'à chaque étape on compare aux deux limites :

$$\frac{\gamma_1}{1-\gamma_0} \quad \text{et} \quad \frac{1-\gamma_1}{\gamma_0}$$

le rapport des fonctions de vraisemblance  $f_1(t_2, t_3 \dots t_n / \delta_1)$  de l'échantillon  $t_2, t_3 \dots$

Cependant les  $t_k$  n'étant pas statistiquement indépendants, (ce qui d'ailleurs n'empêche pas l'application du test de WALD), ces fonctions de vraisemblance apparaissent comme très compliquées à calculer.

L'application d'un théorème général de COX [9] permet une simplification importante, car il indique que  $f_1(t_2, t_3 \dots t_n / \delta_1)$  peut se factoriser sous la forme suivante

$$f_1 = f(t_n / \delta_1) g(t_2 \dots t_{n-1} / t_n)$$

où  $g$  est indépendant de  $\delta_1$ , et où  $f$  est la distribution de  $t_n$  sous  $H_1$ , c'est-à-dire une loi de  $t$  à  $n-1$  d.d.l, non centrée, de paramètre de décentrement  $\delta_1$ .

Finalement tant que

$$\log \frac{\gamma_1}{1 - \gamma_0} < \log \frac{f(t_n/\delta_1)}{f(t_n/\delta_0)} < \log \frac{1 - \gamma_1}{\gamma_0}$$

On continue l'échantillonnage ; on cesse avec la conclusion appropriée dans le cas contraire.

On peut facilement calculer  $f(t_n, \delta)$  et vérifier que le rapport des vraisemblances est

$$\exp \left\{ - \frac{(n - u)^2}{2} (\delta_1^2 - \delta_0^2) \right\} \text{Hh}_{n-1}(-\delta_1 u) / \text{Hh}_{n-1}(-\delta_0 u)$$

On voit donc que cette méthode donne un résultat quasi-identique à la méthode des "Weight Functions" de WALD, la seule différence étant dans le suffixe de la fonction Hh ( $n - 1$  au lieu de  $n - 2$ ).

DAVIES [10] donne quelques tables permettant de tracer les frontières dans le plan  $(n, u)$ . (Ces tables peuvent également servir, avec les modifications appropriées au test décrit en a). Le point le plus intéressant est que ces frontières ne partent pas de  $n = 1$ , mais de valeurs  $n_0$  et  $n_1$  supérieures à 1. Autrement dit, il faut au moins  $\min(n_0, n_1)$  paires pour pouvoir conclure, quelle que soit la valeur des différences observées sur les premières paires.

## 2. Tests fermés

Les seuls travaux concernant les tests fermés t-séquentiels sont ceux de SCHNEIDERMAN et ARMITAGE [16].

L'équivalent fermé des tests ouverts précédents s'obtient en gardant la frontière supérieure, et en remplaçant la frontière inférieure par une verticale d'abscisse  $N^*$ . Il n'y a pas de procédé pour déterminer théoriquement  $N^*$ . On utilise le procédé approximatif suivant du à ARMITAGE.

Si  $N_0$  et  $N_1$  sont les effectifs requis par une analyse classique d'une part quand la variance est connue, et d'autre part quand elle ne l'est pas, si  $N$  est la limite dans le test fermé quand la variance est connue, on admet que

$$\frac{N^*}{N_1} = \frac{N}{N_0}$$

qui permet d'obtenir  $N^*$ .

## C. Le critère est qualitatif

Nous rappelons que nous travaillons alors sur les "préférences".

Les exigences de l'expérimentateur peuvent être résumées dans le tableau général suivant

Pourcentage de préférences en faveur de B	$\theta_0 (H_0)$	$\theta_1 (H_1)$
Conclusion souhaitée	Accepter A	Accepter B
Risques consentis	$\gamma_0$	$\gamma_1$

En attitude explicative, on prendra  $\theta_0 = \frac{1}{2}$ ,  $\theta_1 = \theta$ ;  $\gamma_0 = \alpha$  risque de première espèce,  $\gamma_1 = \beta$  risque de 2ème espèce, les conclusions souhaitées étant dans ce cas, B non significativement supérieur à A, et B supérieur à A.

En attitude pragmatique, cas symétrique, on prendra  $\theta_0$  et  $\theta_1$  symétriques par rapport à 1/2 ( $\theta_0 = 1 - \theta$ ,  $\theta_1 = 1 + \theta$ ),  $\gamma_0 = \gamma_1 = \gamma$ .

Si le point d'équivalence est  $h = 1/2$ , on pourra prendre  $\theta_0 = h + \theta$ ,  $\theta_1 = h - \theta$ ; mais on remarquera alors que la situation n'est pas exactement symétrique par rapport à  $h$ . En particulier la probabilité d'accepter A ou B pour  $h$ , n'est pas exactement 1/2. Cependant comme elle en diffère assez peu, on pourra se contenter de cette approximation.

Le choix de  $\theta_0$  et  $\theta_1$  pour un essai déterminé n'est pas sans poser de réelles difficultés. Nous nous contentons de renvoyer à (18) où ce problème est discuté.

### 1. Test ouvert de WALD

On sait que les frontières sont des droites parallèles de pente commune

$$b = \frac{\log \frac{1 - \theta_0 \cdot \theta_0}{1 - \theta_1 \cdot \theta_1}}{\log \frac{\theta_1 \cdot 1 - \theta_0}{\theta_0 \cdot 1 - \theta_1}}$$

et d'ordonnées à l'origine

$$c_+ = \frac{2 \log \frac{1 - \gamma_1}{\gamma_0}}{\log \frac{\theta_1 \cdot 1 - \theta_0}{\theta_0 \cdot 1 - \theta_1}}$$

$$c_- = \frac{2 \log \frac{\gamma_1}{1 - \gamma_0}}{\log \frac{\theta_1 \cdot 1 - \theta_0}{\theta_0 \cdot 1 - \theta_1}}$$

### 2. Tests fermés

On pourra appliquer les tests vus à propos des critères quantitatifs de variance connue, en assimilant la distribution du nombre de préférences  $\Sigma d$ , à une variable normale. On rappelle que si  $\theta$  est le pourcentage des préférences en faveur de B,  $d$  à pour moyenne  $2\theta - 1$  et pour variance  $4\theta(1 - \theta)$ .

Par exemple l'équivalent du test représenté figure 2 est celui de la figure 7. On voit immédiatement sur la figure qu'on peut remplacer les segments  $M_1N$  et  $M_0N$  par  $NR_1$  et  $NR_0$ , car si le chemin coupe par exemple  $R_1N$ , quelques que soient les observations ultérieures, il coupera nécessairement  $R_1M$  ou  $M_1N$ .

De même, l'équivalent du test I-A-2b est représenté figure 8. La même remarque que précédemment conduit à un test qui est automatiquement en "coin", et très voisin des tests à deux droites concourantes d'ANDERSON.

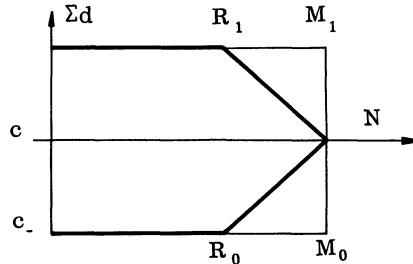


Fig. 7

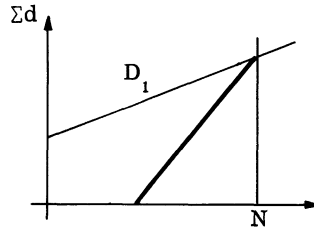


Fig. 8

N se calcule à partir de la formule

$$P(\Delta_1, N) = 1 - \gamma_1$$

dans laquelle

$$\Delta_1 = 2\theta_1 - 1, \sigma_1^2 = 4\theta(1 - \theta)$$

## II - PROBLEMES A TROIS CONCLUSIONS

(Attitude Explicative, situation bilatérale)

Cette attitude correspond, on s'en souvient, au cas où il est grave de conclure à une différence qui n'existerait pas. D'autre part, il est gênant de ne pas déceler une différence qui existerait dans un sens ou dans l'autre.

A. Critère quantitatif, la variance  $\sigma^2$  de  $d = x_B - x_A$  est connue.

Les exigences sont résumées dans le tableau ci-dessous.

$\mu_B - \mu_A$	$-\Delta (H_-)$	$0 (H_0)$	$\Delta (H_+)$
Conclusion souhaitée	$A > B$	A non différent de B	$A < B$
Risques consentis	$\beta$	$\alpha$	$\beta$

### 1. Tests ouverts

Deux procédés sont possibles :

a) le premier du à SOBEL et WALD [21] consiste à effectuer simultanément deux tests  $T_+$  et  $T_-$  à deux conclusions chacun ;  $T_+$  testant  $H_0$  contre  $H_+$  avec les risques  $\alpha/2$  et  $\beta$ , et  $T_-$  testant  $H_0$  contre  $H_-$  avec les mêmes risques, puis de combiner les conclusions de ces deux tests.



b) le deuxième consiste à tester  $H_0$  contre l'hypothèse  $H : |\mu_B - \mu_A| = \Delta$ . Si  $H_0$  est rejetée, le choix entre  $(A < B)$  et  $(A > B)$  se fait par inspection du signe de  $\Sigma d$ .

Nous allons maintenant expliciter ces deux procédés.

1.a. Le test  $T_+$  conduit aux deux droites parallèles  $D_+$  et  $D'_+$ , de pente  $\Delta/2$  et d'ordonnées à l'origine.

$$\frac{\sigma^2}{\Delta} \left( \log \frac{1 - \beta}{\alpha/2}, \log \frac{\beta}{1 - \alpha/2} \right)$$

Le test  $T_-$  conduit aux 2 droites  $D_-$  et  $D'_-$  symétriques des précédentes. (Fig. 9).

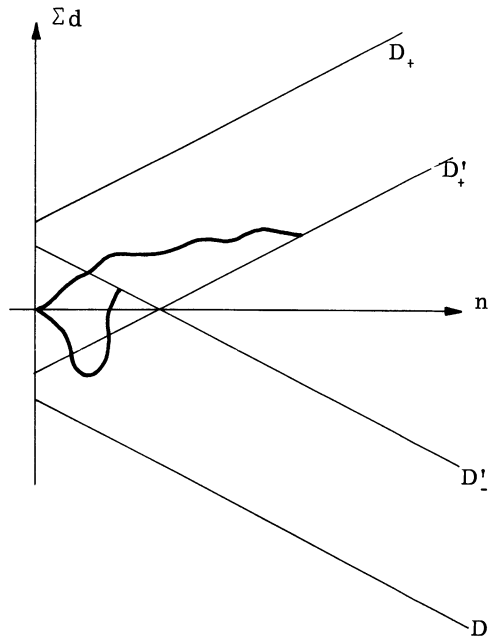


Fig. 9

On effectue les 2 tests  $T_+$  et  $T_-$  simultanément et on arrête chacun d'eux, indépendamment de l'autre, dès que l'une de ses frontières est rencontrée.

Considérant un chemin quelconque, ceci revient à combiner en un test  $T$  unique, les deux tests de la façon suivante :

- on conclut  $A = B$  (différence non significative) dès que le chemin a rencontré les deux droites  $D'_+$  et  $D'_-$ . C'est le cas notamment des chemins représentés sur la figure.

- on conclut  $B > A$  dès que le chemin a rencontré  $D_+$  et  $D'_-$ .

- on conclut  $A > B$  dès que le chemin a rencontré  $D'_+$  et  $D_-$ .

Il s'agit maintenant de montrer que le test  $T$  ainsi construit répond bien aux exigences fixées.

Appelons  $P_+(\delta/T_+)$  et  $P_0(\delta/T_+)$  les probabilités de conclure  $H_+$  et  $H_0$  avec le test  $T_+$ , quand la moyenne de la différence à la valeur  $\delta$ ;  $P_-(\delta/T_-)$  et  $P_0(\delta/T_-)$  d'une part,  $P_+(\delta/T)$ ,  $P_0(\delta/T)$ ,  $P_-(\delta/T)$  de l'autre, ont des définitions analogues.

On a d'abord, puisque  $T$  à la probabilité 1 de se terminer  $P_+(\delta/T) + P_0(\delta/T) + P_-(\delta/T) = 1$ .

Ensuite il est clair d'après la définition du test  $T$ , que  $P_+(\delta/T) = P_+(\delta/T_+)$  et que de même  $P_-(\delta/T) = P_-(\delta/T_-)$ , de sorte que  $P_+(\delta/T_+) + P_0(\delta/T) + P_-(\delta/T_-) = 1$ .

Pour  $\delta = 0$ ,

$$P_+(0/T_+) = P_-(0/T_-) = \alpha/2$$

soit

$$P_0(0/T) = 1 - \alpha$$

Le risque de première espèce du test  $T$  est donc bien  $\alpha$ .

Pour  $\delta = \Delta$

$$P_+(\Delta/T_+) = 1 - \beta,$$

soit

$$P_0(\Delta/T) + P_-(\Delta/T) = \beta$$

la première de ces quantités est la probabilité de conclure  $A = B$ , si  $B > A$ , la deuxième est la probabilité de conclure  $A > B$ , quand  $B > A$ .

La quantité  $\beta$  est donc la somme des risques de deuxième et troisième espèce du test  $T$ .

1.b. Pour tester  $H_0$  contre  $H : |\mu_B - \mu_A| = \Delta$ , on dispose a priori de deux méthodes.

(i). Assigner, la probabilité a priori 1/2 à chacune des hypothèses  $H_+$  et  $H_-$  (méthode des "Weight Functions" de WALD, décrite en I-B-1 a).

La vraisemblance sous  $H$  est alors  $L = 1/2 (L_- + L_+)$  dont le rapport à  $L_0$  doit être comparé à chaque étape aux deux constantes

$$\frac{\beta}{1 - \alpha} \quad \text{et} \quad \frac{1 - \beta}{\alpha}$$

( $L_-$ ,  $L_0$ ,  $L_+$  désignent la vraisemblance sous  $H_-$ ,  $H_0$ ,  $H_+$ ).

On vérifie immédiatement que

$$\frac{L}{L_0} = \frac{1}{2} e^{-n \frac{\Delta^2}{2\sigma^2}} \left( e^{\frac{\Delta}{\sigma^2} \Sigma d} + e^{-\frac{\Delta}{\sigma^2} \Sigma d} \right),$$

On obtient des frontières du type représenté sur la figure 10.

On continue l'essai tant que le point  $(n, \Sigma d)$  tombe à l'intérieur de la zone définie par les frontières. En particulier, il n'est pas impossible (bien que très peu probable) que le chemin saute d'un couloir à l'autre, ce qui n'entraîne pas un arrêt de l'essai.

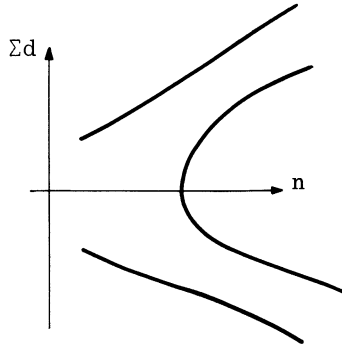


Fig. 10

(ii). Travailler, non plus en termes de  $\Sigma d$ , mais en termes de

$$y_n = \left( \sum_1^n d \right)^2 ;$$

c'est-à-dire remplacer les observations  $d_1, d_2, \dots$  par les observations transformées  $y_1, y_2, \dots$  auxquelles on applique la théorie générale de WALD.

Les  $y_i$  ne sont pas indépendants, mais il est facile de montrer que la distribution  $f$  des  $y$  peut s'écrire.

$$f(y_1, y_2, \dots, y_n/\delta) = g(y_n/\delta) \cdot h(y_1 \dots y_{n-1}/y_n)$$

où  $h$  est indépendant du paramètre  $\delta$ .

Le test revient donc à comparer à

$$\frac{\beta}{1 - \alpha} \quad \text{et} \quad \frac{1 - \beta}{\alpha}$$

le rapport

$$g(y_n/\Delta) / g(y_n/0)$$

$y_n$  étant le carré d'une variable normale de variance  $n\sigma^2$  et de moyenne  $n\delta$ ,  $\frac{y_n}{n\sigma^2}$  est sous  $H_0$  un  $\chi^2$  à 1 degré de liberté, et sous  $H$  un  $\chi^2$  non centré avec le paramètre de décentrement  $\frac{\sqrt{n}\Delta}{\sigma}$ .

Il est alors facile de vérifier que  $g(y_n/\Delta)/g(y_n/0)$  est égal à la quantité  $L/L_0$  du paragraphe (i) précédent.

Ainsi donc les procédés (i) et (ii) conduisent exactement au même test : le deuxième paraît toutefois d'une présentation plus intéressante, les "Weight Functions" ayant toujours un aspect un peu arbitraire.

Dans ce test,  $\beta$  est la probabilité de conclure  $H_0$  quand  $|\mu_B - \mu_A| = \Delta$ , c'est donc strictement le risque de deuxième espèce, et il a une signification légèrement différente du  $\beta$  du test (a).

### 1.c. Comparaison des tests (a) et (b).

Les frontières extérieures sont extrêmement voisines : en effet la supérieure est définie par

$$\frac{L_+}{L_0} = \frac{1 - \beta}{\alpha/2} \text{ pour le test (a)}$$

et par

$$\frac{1}{2} \frac{L_- + L_+}{L_0} = \frac{1 - \beta}{\alpha} \text{ pour le test (b)}$$

Comme au voisinage de cette frontière,  $L_-$  est négligeable devant  $L_+$ , les deux définitions se confondent.

La différence est plus importante pour les frontières intérieures.

Quant aux propriétés des deux tests, en particulier, le nombre moyen d'observations, elles semblent très proches : aussi utilise-t-on presque exclusivement le premier, car il a l'avantage pratique considérable de conduire à des frontières linéaires, immédiatement calculables.

### 2. Tests fermés

Seules ont été considérées les variantes fermées du test II-A-1a.

#### a) Tests "limités" d'ARMITAGE [4].

On conserve les frontières linéaires extérieures  $D_+$  et  $D_-$ , et on remplace les droites intérieures par une verticale d'abscisse  $N$ , calculée pour que les risques voulus soient préservés (fig. 11).

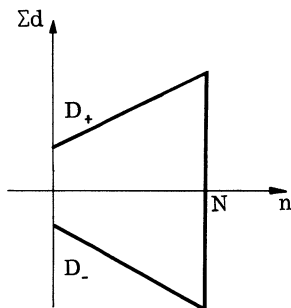


Fig. 11

Ceci revient à combiner par le procédé de WALD - SOBEL 2 tests à deux réponses, définis en I-A-2b. Il suffit donc d'utiliser les formules qui sont données à ce paragraphe, en faisant

$$\gamma_0 = \frac{\alpha}{2}, \gamma_1 = \beta, \Delta_1 = \Delta_0, \Delta = 0$$

Le procédé cependant n'est en toute rigueur qu'approché, car l'égalité  $P_+(\delta/T) = P_+(\delta/T_+)$  n'est pas strictement vérifiée (elle revient en fait à négliger des chemins qui, commenceraient par couper  $D_-$  avant d'atteindre  $D_+$ ).

ARMITAGE montre que l'approximation est excellente, du moins dans les cas pratiques.

b) Tests "en coin" (SCHNEIDERMAN et ARMITAGE [16].

Deux variantes sont possibles.

(i). Combiner par le procédé de WALD et SOBEL, deux tests "en coin" à deux réponses décrits en I.A.2.C., ce qui donne des frontières telles que celles représentées sur la figure 12.

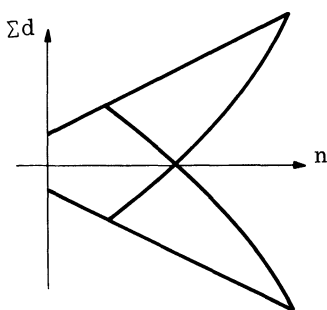


Fig. 12

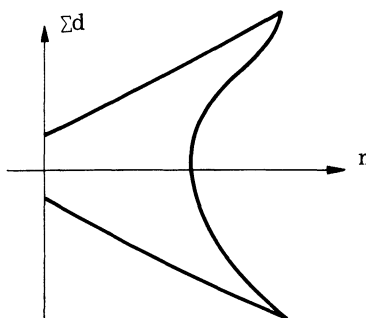


Fig. 13

(ii). Modifier le test représenté fig. 11 pour obtenir le schéma de la fig. 13. La frontière intérieure étant obtenue par un procédé analogue à celui du paragraphe I-A-2c.

B. Critère quantitatif, la variance  $\sigma^2$  de  $d = x_B - x_A$  est inconnue.

Les exigences sont résumées dans le tableau. On se souvient qu'elles sont exprimées par rapport à  $\sigma$ .

$ \mu_B - \mu_B $	$-\delta\sigma(H_-)$	$0(H_0)$	$\delta\sigma(H_+)$
Conclusion souhaitée	A > B	A non différent de B	A < B
Risques consentis	$\beta$	$\alpha$	$\beta$

### 1. Tests ouverts

A priori quatre tests sont possibles. En effet on a décrit deux tests ouverts t séquentiels à deux réponses : les tests de WALD et de RUSHTON-BARNARD, d'ailleurs très voisins ; et on a vu qu'il y avait deux façons de construire un test à trois réponses : le procédé de SOBEL-WALD, et le test de  $H_0$  contre  $H : |\mu_B - \mu_A| = \delta\sigma$ . On peut donc résumer :

Test t à 2 réponses	SOBEL-WALD	$H_0/H$
WALD	(a)	(c)
RUSHTON-BARNARD	(b)	(d)

Il n'est pas nécessaire de redonner ici les formules définissant ces 4 tests. On se reportera aux pages qui précèdent. Les tests (a) et (b) sont représentés sur la figure 14 et les tests (c) et (d) sur la figure 15. On rappelle que le chemin représente la variation de  $u$  en fonction de  $n$

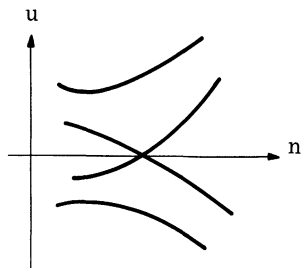


Fig. 14

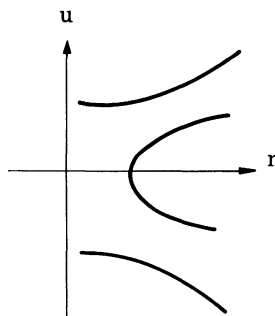


Fig. 15

Au point de vue pratique, on a le choix entre les tests (b) et (d).

Pour le test (b) on dispose des tables de DAVIES, déjà citées en I-B-1 b ; pour le test (d), des tables extrêmement détaillées du National Bureau of Standards (13).

C'est à cause de la richesse de ces tables que le test d est le plus utilisé. De plus RUSHTON [15] donne un procédé pour obtenir les frontières au delà des situations couvertes par les tables. Ces dernières sont relatives à la statistique  $u^2 = (\Sigma d)^2 / \Sigma d^2$ , ce qui donne les frontières de la figure 16. On stoppe en concluant : différence non significative si I est traversée ; différence significative si S est traversée, le signe de cette différence étant donné par le signe de  $\Sigma d$  au moment où on stoppe.

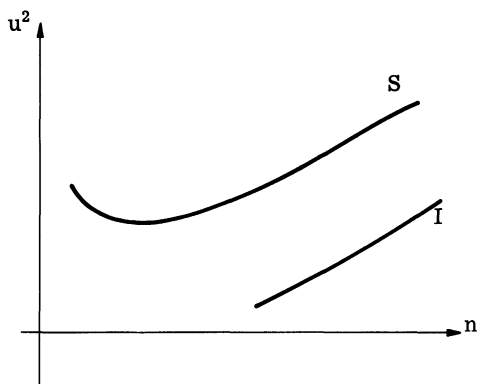


Fig. 16

On remarquera que pour  $n$  grand, S et I sont pratiquement des droites Ceci est facile à démontrer. On rappelle également que S et I ne démarrent pas à  $n = 0$ , mais qu'il faut, quelles que soient les premières différences observées, un nombre minimum d'observations pour pouvoir conclure, et que S commence par décroître.

## 2. Tests fermés

Ils n'ont semble-t-il été considérés que par SCHNEIDERMAN et ARMITAGE [17].

Dans les tests "limités", on garde les frontières extérieures des tests du paragraphe précédent et on remplace les frontières intérieures par une verticale d'abscisse  $N$  (fig. 17) ; pour les tests (c) et (d) où on a l'habitude comme on l'a dit de travailler avec  $u^2$  plutôt qu'avec  $u$ , on obtient la figure 18.

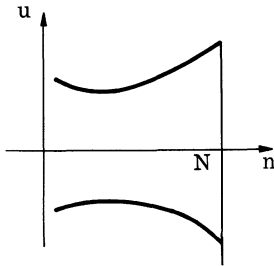


Fig. 17

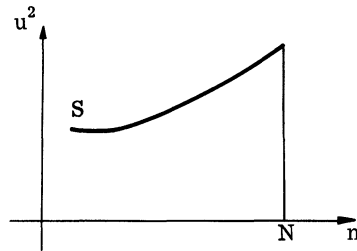


Fig. 18

Pour déterminer  $N^*$ , on utilisera la conjecture d'ARMITAGE (Voir I-B-1)

### C. Le critère est qualitatif

Les exigences se formulent ainsi

Pourcentage de préférences en fonction de B	$1 - \theta (H_1)$	$1/2 (H_0)$	$\theta (H_1)$
Conclusion souhaitée	$A > B$	A non différent de B	$A < B$
Risques consentis	$\beta$	$\alpha$	$\beta$

#### 1. Tests ouverts

On peut appliquer les raisonnements faits à propos des critères quantitatifs.

a. Procédé de SOBEL-WALD [21] : on obtient deux couples symétriques de droites parallèles.

b. Test de  $H_0 : 1/2$  contre  $H : \theta$  ou  $1-\theta$  par la méthode des "Weight Functions".

Alors si  $x$  et  $y$  désignent les nombres de préférences en faveur de A et B ( $x + y = n$ ), et  $\Sigma d$  la différence  $y - x$ , la vraisemblance sous  $H_0$  est  $L_0 = \binom{n}{y} \left(\frac{1}{2}\right)^n$  et sous H elle est

$$L = \frac{1}{2} \binom{n}{y} \left( \theta^y (1 - \theta)^x + \theta^x (1 - \theta)^y \right),$$

de sorte que

$$\frac{L}{L_0} = 2^{n-1} \left( \theta \frac{n+\Sigma d}{2} (1-\theta)^{\frac{n-\Sigma d}{2}} + \theta \frac{n-\Sigma d}{2} (1-\theta)^{\frac{n+\Sigma d}{2}} \right)$$

qui est à comparer aux 2 constantes

$$\frac{\beta}{1-\alpha} \quad \text{et} \quad \frac{1-\beta}{\alpha}$$

on obtient ainsi la figure 19. Ce test n'est pas utilisé, le premier étant immédiat à construire.

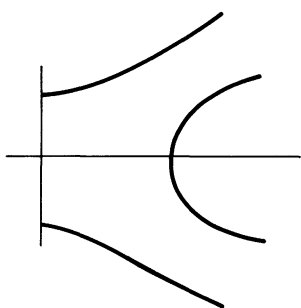


Fig. 19

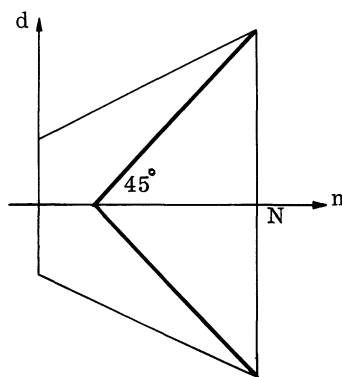


Fig. 20

## 2. Tests fermés

a. Seule a été considérée la version fermée du test II.C.1.a précédent. On garde les droites extérieures et on remplace les droites intérieures par une verticale d'abscisse N. (Fig. 20).

N se calcule en assimilant  $\Sigma d$  à une variable normale, d, ayant pour moyenne  $2\theta - 1$  et pour variance  $4\theta(1-\theta)$  ;

On notera que ce test "limité à N" se transforme automatiquement en "test en coin".

ARMITAGE [5] donne les valeurs de N pour différentes valeurs de  $\theta$  et des risques  $\alpha$  et  $\beta$ , obtenues à partir de l'approximation par la loi normale.

Cependant quand N est petit, l'approximation est insuffisante, ARMITAGE donne alors en calculant les probabilités exactes de rencontrer les frontières, de schémas qui conduisent à des risques très voisins de ceux fixés.

b. BROSS [6], SPICER [22] ont donné quelques tests construits plus ou moins empiriquement.

## III - RETOUR SUR LE TEST UNILATERAL EN ATTITUDE EXPLICATIVE

Supposons qu'on travaille avec des effectifs fixés à l'avance et que le critère soit quantitatif ; dans le test unilatéral on ne conclut à la significa-



tion que si  $m_B$  dépasse  $m_A$  d'une certaine quantité ; lorsque  $m_B$  est inférieure à  $m_A$  même très largement, on ne tient aucun compte de cet écart, l'hypothèse  $B < A$  étant soit exclue, soit inintéressante. Cependant une telle politique en attitude explicative est peut réaliste : en effet quel chercheur voyant une différence très significative dans le sens inattendu, décidera de l'ignorer ? On peut proposer la démarche suivante dans le cas d'effectifs fixés (18) : déterminer l'effectif à partir des formules établies pour le test unilatéral ; conclure  $B > A$  si  $m_B$  dépasse  $m_A$  au seuil fixé pour un test unilatéral par exemple ; mais ne conclure  $A > B$  que si  $m$  dépasse très significativement  $m_B$ , par exemple si  $\alpha = 0.001$ . Ceci revient à avoir un risque total de première espèce pratiquement inchangé ; en contrepartie la puissance du test est faible pour des valeurs négatives de  $m_B - m_A$ .

On peut étendre ce procédé à l'analyse séquentielle : nous raisonnons dans le cas où le test est celui de WALD ;  $D_1$  et  $D_0$  sont les droites habituelles c'est-à-dire que

$$\text{Pr (de couper } D_1/\Delta) = 1 - \beta$$

$$\text{Pr (de couper } D_1/0) = \alpha \text{ (par exemple 0.05).}$$

Déterminons sur  $D_0$  une zone  $D'_0$  telle que

$$\text{Pr (de couper } D'_0/0) = \alpha' \text{ (très faible, par exemple 0.001).}$$

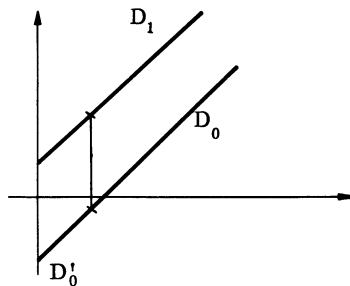


Fig. 21

Nous concluons  $B > A$  si le chemin coupe  $D_1$ ,  $A < B$  s'il coupe  $D'_0$  ;  $A = B$  s'il coupe le reste de la frontière  $D_0$ .

La détermination de  $D'_0$  présente des difficultés plus ou moins grandes selon le test utilisé : dans le cas où les frontières sont des droites (notamment dans le cas de variables normales de variance connue) on peut, théoriquement se servir des formules données par ANDERSON [1] ; si ce ne sont pas des droites (tests t-séquentiels par exemple), il ne semble pas y avoir à l'heure actuelle de procédé disponible.

#### IV - DIFFICULTES SOULEVEES PAR L'APPLICATION DE LA METHODE SEQUENTIELLE AUX ESSAIS THERAPEUTIQUES

1/ On présente très généralement l'analyse séquentielle pour des raisons d'éthique (nécessité morale de stopper un essai dès qu'on a l'assurance suffisante qu'un des traitements est meilleur que l'autre) ; mais d'un autre côté nous avons dit dans l'introduction qu'elle apparaissait être le plus utile quand la réponse aux traitements est obtenue dans un temps bref, ce qui le

plus souvent est signe que la maladie n'est pas très grave (maladies aiguës ; crises de maladies chroniques ...), auquel cas le problème éthique perd un peu de son acuité.

La question qui se pose alors est la suivante : est-il possible d'utiliser l'analyse séquentielle quand la réponse est longue à venir, par exemple quand il s'agit de comparer les taux de survie, ou l'état des sujets un an après l'administration des traitements.

Examinons tout d'abord ce dernier cas : supposons que l'essai commence au temps 0 et que les malades y entrent régulièrement au rythme de 200 paires par an. Les premières réponses commencent à parvenir au bout d'un an, moment à partir duquel on peut commencer à tracer le chemin. Admettons qu'au bout de 100 paires, le chemin coupe une des frontières, par exemple celle correspondant à la conclusion "A meilleur que B". Cette conclusion qui est basée sur les 100 premières paires entrées dans l'essai de 0 à T, sera obtenue au bout de 1 an et demi, en T'. Supposons qu'on stoppe alors l'essai et qu'on adopte définitivement le traitement A. Mais on n'a pas encore eu la réponse des sujets traités entre T et T', qui sont au nombre de 200 paires. Elles vont continuer à nous parvenir jusqu'en T'', et il est nécessaire d'en tenir compte d'une façon ou d'une autre.

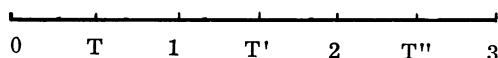


Fig. 22

Un cas particulier relativement simple est celui où après avoir stoppé en T' par l'adoption de A, on décide de refaire une analyse finale en T'' quand toutes les réponses ont été obtenues.

ANDERSON [3] aborde ce problème mais sous des conditions restrictives (le nombre supplémentaire de paires est fixé à l'avance ; le test séquentiel est celui de WALD).

Le cas qui vient d'être considéré est simple mais pas très réaliste car les réponses arrivent séquentiellement et il n'est pas raisonnable de vérifier seulement en T'' le résultat obtenu en T', alors qu'on pouvait envisager de le faire tout au long de T', T''. Ce problème n'a pas de solution à l'heure actuelle.

Un cas encore plus difficile est celui où la réponse serait obtenue après un temps non plus fixe (un an, comme dans l'exemple précédent) mais aléatoire (durée de survie).

2/ D'autre part, l'analyse séquentielle me rend-elle pas plus aiguës les difficultés éthiques qu'elle devait supprimer ?

Supposons que le chemin ait l'allure ci-contre (fig. 23) : la signification n'est pas encore atteinte mais il semble bien qu'on va y arriver très prochainement. L'éthique n'impose-t-elle pas de stopper en adoptant B ? Cependant si on stoppait à ce moment les risques ne seraient plus ceux fixés.

On pourrait alors envisager une autre procédure. Au lieu de donner, comme d'habitude, les 2 traitements A et B aux deux éléments de chaque paire, on donne à chaque malade, qui entre dans l'essai, le traitement qui à ce moment "semble" être le meilleur, compte tenu de tous les résultats antérieures. Chaque malade reçoit ainsi la "meilleure thérapeutique" du moment.

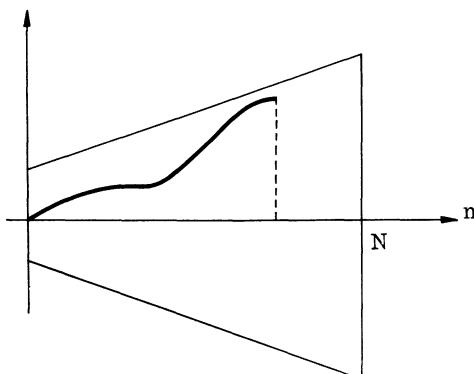


Fig. 23

C'est certainement une forme très séduisante d'essai thérapeutique, à la fois pour le médecin qui à chaque instant agit au mieux, compte tenu de ses connaissances, et pour le statisticien à qui elle pose un certain nombre de difficiles problèmes théoriques, à peine abordés.

c) De plus, certains des schémas présentés plus haut peuvent être difficilement acceptables d'un point de vue éthique : dans la figure 24 dès que le chemin pénètre dans la zone hachurée, on est sûr que A ne peut être meilleur que B. A-t-on le droit de continuer l'essai, ou doit-on dès ce moment le stopper, en adoptant B, qui ne peut être que supérieur ou égal ?

En fait ce schéma convient pour des problèmes de "recherche" et l'on sait les difficultés que l'on rencontre à faire de la "recherche" sur l'homme.

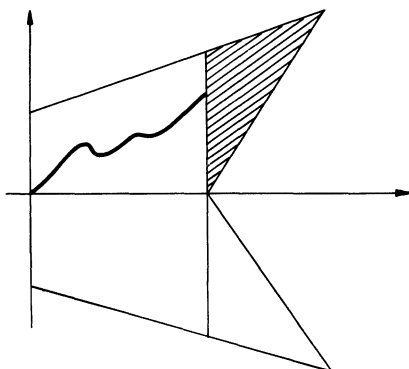


Fig. 24

d) L'analyse séquentielle présente enfin d'autres difficultés, que nous soulignons brièvement. Les frontières sont établies pour un critère déterminé. Si on veut comparer tout un ensemble de critères, la meilleure façon de procéder est de les combiner en un indice unique, pour lequel on trace les frontières adéquates. Si cette solution peut parfaitement convenir à des problèmes pragmatiques, il peut ne pas en être du tout de même pour des problèmes de recherche, où l'on a à comparer isolément une série de critères.

Il est de même difficile de comparer les traitements sur des sous-groupes (hommes et femmes par exemple) une fois obtenu un résultat sur l'ensemble.

## BIBLIOGRAPHIE

- [1] ANDERSON T.W. - A modification of the sequential probability ratio test to reduce the sample size. *Annals of Mathematical Statistics* Vol.31, 31, n° 1, March (1960).
- [2] ANDERSON T.W. - Sequential analysis with delayed observations. *Journal of the American Statistical Association*, December Vol. 59 pp. 1006-1015 (1964).
- [3] ANSCOMBE F.J. - Sequential medical trials. *Journal of the American Statistical Association*, June, Vol. 58, pp. 365-383 (1963).
- [4] ARMITAGE P. - Restricted Sequential procedures. *Biometrika* 44. 9 - 26 - 1957.
- [5] ARMITAGE P. - Sequential medical trials, Blackwell (1960).
- [6] BROSS I. - Sequential medical plans. *Biometrics* 8. 188 (1952).
- [7] COLTON T. - A model for selecting one of two medical treatments. *Bulletin de l'Institut International de Statistique* 33e Session, Paris (1961).
- [8] CORNFIELD J. - Sequential trials, sequential analysis and the Likelihood principle. *The Amer. Statist.* Vol. 20. N° 2 p. 18-23 (1966).
- [9] COX D.R. - Sequential tests for composite hypotheses. *Proceedings of the Cambridge Philosophical Society*, Vol. 48. Part. 2, pp. 290-299, (1952).
- [10] DAVIES O.L. (Edit.) - Désign and analysis of industrial experiments . Oliver and Boyd (1954).
- [11] FELLER W. - An introduction to probability theory and its applications . Vol. I Wiley (1957).
- [12] GENIZI A. - On the performance of the truncated sequential probability ratio test. *Journal of the American Statistical Association*. December Vol. 60, pp. 979-984 (1965).
- [13] NATIONAL BUREAU OF STANDARDS - Tables to facilitate sequential t-tests. Washington U.S. Dep. of Commerce (1951).
- [14] RUSHTON S. - On a sequential t-test. *Biometrika* 37 ; 326. (1950).
- [15] RUSHTON S. - On a two-sided sequential t-test. *Biometrika*, Vol. 39, Parts 3 and 4, December (1952).
- [16] SCHNEIDERMAN M.A. and ARMITAGE P. - A family of closed sequential procedures. *Biometrika* 49, 1 and 2, p. 41 (1962).
- [17] SCHNEIDERMAN M.A. and ARMITAGE P. - Closed sequential t-tests. *Biom.* 49. 359 (1962).
- [18] SCHWARTZ D. et FLAMANT R. - Les essais Thérapeutiques (à paraître).
- [19] SCHWARTZ D. et LELLOUCH J. - Attitude de recherche ou attitude pragmatique dans la formulation d'un problème, avec référence particulière aux essais thérapeutiques. *Revue de Statistique Appliquée*. Vol. XIII, N° 4 (1965).
- [20] SCHWARTZ D. et LELLOUCH J. - Explanatory and pragmatic attitudes in therapeutical trials. *J. Chron. Dis.* Vol. 20, pp. 637-648 (1967).

- [21] SOBEL M. and WALD A. - A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.* 20 ; 502 (1949).
- [22] SPICER C.C. - Some new closed sequential designs for clinical trials. *Biometrics* 18 ; 203 (1962).
- [23] WALD A. - Sequential Analysis. Wiley (1947).